

This article was downloaded by: [USDA Natl Agricultul Lib]

On: 24 December 2009

Access details: Access Details: [subscription number 731827127]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Animal Biotechnology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713597228>

### CHARACTERIZATION OF A NORMALIZED CDNA LIBRARY FROM BOVINE INTESTINAL MUSCLE AND EPITHELIAL TISSUES

R. G. Baumann <sup>a</sup>; R. L. Baldwin VI <sup>a</sup>; C. P. Van Tassell <sup>a</sup>; T. S. Sonstegard <sup>a</sup>; L. K. Matukumalli <sup>ab</sup>

<sup>a</sup> USDA, ARS, ANRI, Bovine Functional Genomics Laboratory, Beltsville Agricultural Research Center, Beltsville, Maryland, USA <sup>b</sup> Bioinformatics and Computational Biology, George Mason University, Manassas, Virginia, USA

**To cite this Article** Baumann, R. G., Baldwin VI, R. L., Van Tassell, C. P., Sonstegard, T. S. and Matukumalli, L. K. (2005) 'CHARACTERIZATION OF A NORMALIZED CDNA LIBRARY FROM BOVINE INTESTINAL MUSCLE AND EPITHELIAL TISSUES', *Animal Biotechnology*, 16: 1, 17 – 29

**To link to this Article:** DOI: 10.1081/ABIO-200053398

**URL:** <http://dx.doi.org/10.1081/ABIO-200053398>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## CHARACTERIZATION OF A NORMALIZED cDNA LIBRARY FROM BOVINE INTESTINAL MUSCLE AND EPITHELIAL TISSUES

**R. G. Baumann, R. L. Baldwin VI, C. P. Van Tassell,  
and T. S. Sonstegard**

*USDA, ARS, ANRI, Bovine Functional Genomics Laboratory, Beltsville  
Agricultural Research Center, Beltsville, Maryland, USA*

**L. K. Matukumalli**

*USDA, ARS, ANRI, Bovine Functional Genomics Laboratory,  
Beltsville Agricultural Research Center, Beltsville, Maryland, USA  
Bioinformatics and Computational Biology, George Mason University,  
Manassas, Virginia, USA*

*Tissue-specific cDNA library sequences (expressed sequence tags, or EST) yield a detailed snapshot of gene expression and are useful in developing second-generation molecular resources (i.e., microarrays) for gene expression profiling. The objective of this study was to develop and characterize an intestine-specific cDNA library to examine the transcriptome of the bovine gut and identify expressed genes that influence ruminant nutrition and health. We describe BARC-8BOV, a normalized cDNA library developed from mRNA isolated from four distinct intestinal locations (duodenal, jejunal and ileal small intestine, colon) of Holstein dairy cattle resulting in 19,110 5'-EST deposited into the NCBI GenBank EST database. Assembly and clustering of these 19,110 clone sequences yielded 11,208 unique elements (3,419 contigs and 7,789 singletons) with an average length of 695 base pairs. Analysis strongly suggests normalization and tissue pooling were effective at increasing the discovery rate of new bovine sequence. A total of 1,123 sequence elements not previously identified in cattle, but with similarity to known genes in other animal species, were identified and shown to be involved in numerous critical biological processes. An additional 745 transcripts were not previously represented as EST in nucleotide or protein databases, and further analysis of these could lead to the identification of gut-specific transcript variants of known genes or potentially the discovery of novel bovine genes. Of the 11,208 assembled sequences, 11,034, or 98.4%, match sequences present in the bovine DNA trace archive at NCBI, and add to a bovine EST database previously lacking significant gut tissue representation. Ultimately, these data will also contribute in efforts to annotate the bovine genome.*

## INTRODUCTION

The inherent limitation to efficiently convert plant carbon and nitrogen into animal products with minimal environmental impact is the paramount concern

Address correspondence to Richard Baumann, Ph.D., USDA; ARS; ANRI, Bovine Functional Genomics Laboratory, Bldg. 162, Rm. 211, BARC-East, Beltsville, MD 20705-2350, USA. E-mail: rbaumann@anri.barc.usda.gov

facing livestock producers. Visceral tissues (gut and liver) are the first barrier of nutrient absorption, and account for approximately 50% of whole body energy expenditure and protein turnover, and thus have a considerable impact on the partitioning of nutrients between tissue accretion, milk production, and excretion losses (1). Therefore, identification of physiological mechanisms regulating nutrient absorption and assimilation is essential for improving animal nutrient economy. A first step in delineating biological mechanisms that have not yet been well characterized by classical physiological study is to develop a cDNA library specifically from these tissues. Indeed a growing number of molecular techniques, used to identify the biological function of genes, rely on the generation and characterization of transcript information (2, 3). A cDNA library specifically from dairy cow visceral tissue contributes by identifying genes involved in major cellular processes, such as transporters, signaling receptors, growth factors, transcription or translation factors, and other proteins that affect ruminant visceral nutrient metabolism.

Normalized cDNA libraries from pooled ruminant tissues have proven to be an excellent source for bovine EST relevant to animal production and health (4–7). In this paper, the construction and analysis of a normalized cDNA library from bovine intestinal muscle and epithelial tissue is described. To increase gene discovery, multiple distinct regions of the intestine (duodenum, jejunum, ileum, colon) were pooled, and intestinal samples were taken from a lactating dairy cow and a pre-suckling neonatal calf. In doing this, the library captures genes expressed in mature gut, as well as newborn prior to gut closure, and unaffected by luminal nutrients. Our aim is to develop this library resource for use in second-generation research tools, particularly microarrays, and perform experiments designed to understand how gene expression patterns and control are coupled with cellular, tissue, and whole animal nutrient use profiles. Developing this resource therefore expands classical approaches to understanding bovine nutrition by complementing ongoing whole animal experiments with gene level information necessary to characterize the molecular controls of producing ruminants (8). Ultimately this work will facilitate development of new nutritional management and selection strategies to optimize animal production.

## MATERIALS AND METHODS

### Abbreviations

BARC—Beltsville Agricultural Research Center; BtGI—*Bos taurus* Gene Index; EST—expressed sequence tags; NCBI—National Center for Biotechnology Information; GO—gene ontology.

### BARC-8BOV Library Construction and Normalization

One multiparous mid-lactation Holstein dairy cow and one pre-suckling Holstein neonatal calf were processed at the Beltsville Agricultural Research Center (BARC) abattoir. Four tissue samples were collected, each 10-cm (~70 g) segments taken from the lactating dairy cow immediately after sacrifice from: (1) proximal duodenum (1 m post-pylorus), (2) jejunum (~10 m from the pylorus), (3) ileum (1 m prior to the ileocecal junction), and (4) colon. No effort was made to separate

epithelial cells from the intestinal segments prior to processing. For representation of EST in neonatal gut tissue unaffected by luminal nutrients, three analogous small intestinal samples from: (1) proximal duodenum (1 m post-pylorus), (2) jejunum (middle of intestine), and (3) ileum (1 m prior to the ileocecal junction) were taken from a one-day-old male neonatal calf and combined at tissue level for one-fifth representation. All tissues were rinsed briefly in ice-cold physiological saline to remove digesta, weighed, and flash frozen in liquid nitrogen. Total RNA extraction, mRNA purification, primary library construction, and normalization were purchased as a service from the Life Technologies division of Invitrogen (Carlsbad, CA). Poly-A-purified mRNA (10 µg) from each tissue sample was pooled in equimolar amounts (50 µg total mRNA) for cDNA synthesis. While pooling was done to increase overall EST, it does introduce the limitation of not knowing from what sample or stage each EST is derived. For this reason, original RNA samples from each source were also preserved. Generated cDNAs were modified by addition of polyT-*NotI* restriction site containing linker and unidirectional cloning was performed by ligation into *NotI/EcoRV* cut vector pCMV-Sport6.1 and transformation into DH10B-TonA cells (T1 and T5 phage resistant). The primary library (defined as NN-8BOV) was normalized using a proprietary self-subtracting method (Invitrogen) using Gene2/*ExoIII* methodology and optimized Cot hybridization conditions to reduce the abundance of high copy number transcripts, similar to that described by Smith and colleagues (2001). The resulting normalized library was defined as BARC-8BOV.

### Clone Processing, EST Sequencing, and Sequence Analysis

A sample of BARC-8BOV library was sent to BACPAC Resources (Oakland, CA), for picking, arraying, and archiving glycerol stocks of 55,296 individual colonies in SOC medium with 50 µg/mL ampicillin and 20% glycerol on 144 384-well plates. For processing, glycerol stock plates were thawed once, pin-stamped into four 96-well growth plates containing 150 µL of LB media with ampicillin (75 µg/mL), and grown for 5–7 h at 37°C with slight agitation. Plates were then restamped into four 96-well overnight growth plates containing 1.2 mL LB with ampicillin, and shaken vigorously overnight at 37°C. Plasmid DNA preparations were performed using Eppendorf 96V Perfect Prep kits (Brinkmann, Westbury, NY) with minor changes to protocol, eluting the purified plasmid DNA into 75 µL of molecular biology grade water. Plasmids (2 µL samples) were transferred to 384-well plates, and 5'-end sequencing reactions (5 µL total volume) were performed using primer SP6E: 5'-AGGCCTATTTAGGTGACACTATAGAAC-3' on ABI 3700 and 3730 DNA sequencers as described (9), using commercially available reagents (Big Dye v1.1 and 2.0) from Applied Biosystems (Foster City, CA). Trace files were processed for Phred quality score >18 (10) and vector sequence information removal using EST-PAGE (11). In total, 23,808 sequencing reactions were performed on clones in the first 62 (384-well) plates of the 8BOV library (144 plates).

All EST were assembled and clustered into unique elements which were processed by RepeatMasker (A. Smit, P. Green, and colleagues, <http://repeatmasker.genome.washington.edu>) and compared with a number of databases using BLAST

(12). Sequences were compared against Release 9 of the non-redundant *Bos taurus* Gene Index (BtGI) at the Institute for Genomic Research (TIGR; [http://www.tigr.org/tigr-scripts/tgi/T\\_index.cgi?species=cattle](http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=cattle)), the human genome sequence assembly, human Refseq, the NR-non-redundant nucleotide, NR-non-redundant protein, and SwissProt databases from the National Center for Biotechnology Information, NCBI (<http://www.ncbi.nlm.nih.gov/>). A cutoff value of  $E10^{-10}$  was used to score a sequence as having a significant match. Sequence elements were also compared against the bovine DNA trace archive at NCBI on August 4, 2004 (<http://www.ncbi.nlm.nih.gov/Traces>).

### Estimation of Normalization Efficiencies

Sequence analysis and clustering was performed on 384 clones from the non-normalized 8BOV library, NN-8BOV. The identities and frequencies of contigs of ~1% frequency or greater (three or more sequences of 334 total) derived from this analysis were determined by BLAST analysis to BtGI Release 10 and compared to the frequencies and identities of the most prevalent contig sequences in our 8BOV normalized library. For 14 NN-8BOV contigs comprising only two sequences (2/334 or of a frequency of 0.6% in NN-8BOV), a 146/14, or 10.4 sequence average was calculated for their prevalence in 8BOV corresponding to a frequency of 0.054 in 8BOV.

### Summarization of 8BOV Gene Ontology (GO) Terms

Of all unique elements in the BARC-8BOV library, and the previously described and normalized BARC-5BOV bovine mammary library (6), we identified those with BLAST hits with the human RefSeq database which had at least one GO annotation term listed under the NCBI LocusLink gene ontology list (<http://www.geneontology.org/GO.doc.html>). A complete list of the first three levels of GO annotation and final level ontological term was obtained for these sequences. For this analysis, all third-level GO terms for each main organizing principle of GO were summarized into a final list. For both libraries these terms were reduced identically; biological process terms were combined from a total of 49 terms to 14, cellular component terms from 24 to 6, and molecular function terms from 108 to 21. It is common for a specific gene to have numerous GO annotations associated with it. In many cases, however, duplicate and identical level 3 GO terms (derived via diverse GO level 1 and GO level 2 annotations and present under any particular principle group) could be eliminated for this summary. Further summarization was achieved by grouping small sets of more specific GO level 3 terminologies into more general and inclusive GO level 3 terms. To prevent the redundancy of gene counts occurring within a summarized GO level 3 category, or to reduce the significance of such an occurrence, every attempt was made to minimize this by maintaining prominent categories without reduction, grouping together only small, distinct, and unique groups together for an accurate representation of a count, or by elimination of like GO level 3 terms from the summarized table.

### **Categorization of Newly Identified Physiologically Relevant 8BOV Sequences**

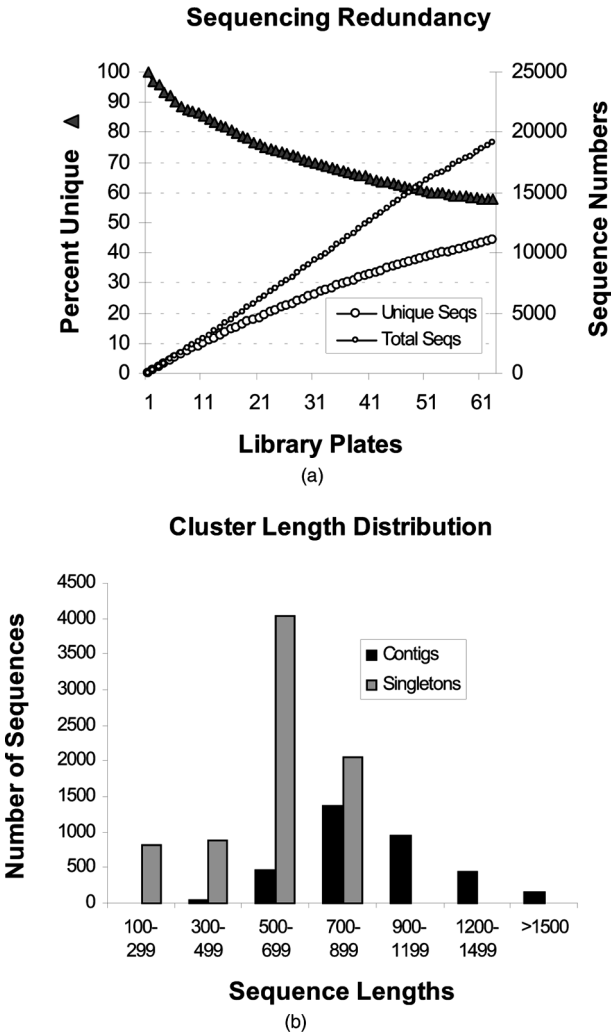
A list of available level 1, 2, 3 and final GO biological process principle terms was generated for new bovine sequences considered physiologically relevant by comparison with, and their presence in, other available databases. From this list, the final level GO terminology, and preceding levels if necessary, was scored individually to group the gene-encoding sequences into the putative biological processes in which they are involved. Sequences were grouped into major categories that highlighted their primary role in molecular metabolism, transport, or cellular macromolecular processes in which they were involved. Many genes were able to be placed under a single category only, and although there were gene sequences scored into more than one category by necessity (due to their diverse biological functions or activities), no sequence is represented in any one category more than once.

## **RESULTS AND DISCUSSION**

We have constructed a normalized cDNA library from whole intestinal tissues of dairy cattle that has resulted in a total of 19,110 5'-EST sequences deposited in the GenBank EST database. An analysis of the sequencing efficiency throughout this process is presented graphically in Figure 1 A. The sequencing success rate was typically very high and even at the completion of 62 plates, the percentage of unique sequences (Figure 1 A, left vertical axis) derived per plate remained high (~58%). Our goal was to derive greater than 10,000 unique elements (singletons plus contigs), which we achieved after 54 plates. We processed additional plates while the discovery rate remained higher than 60%, increasing our total EST and total unique sequence element count over 11,000. Size distribution of all the 8BOV library elements is presented in Figure 1B. Average singleton sequence length was 600 bases, average contig length was 940 bases, giving an average element size of 695 bases. This size distribution analysis of all sequence elements shows that a majority (~80%) are between 500 and 1200 bases long (Figure 1B).

An assembly and complete BLAST analysis was completed for the 19,100 5'-EST derived from BARC-8BOV, and a summary of results is presented in Table 1. Overall, the sequencing success rate was 80%. Clustering of sequences resulted in 11,208 unique elements; a total of 7,789 singletons and 3,419 contigs consisting of 11,321 sequences, giving an average of 3.3 sequences per contig. BLAST comparisons were performed against the current databases at TIGR (BtGI, Release 9) and NCBI, prior to sequence input from the BARC-8BOV library. Eighty three percent of the sequence elements in 8BOV matched those already present in BtGI; however, inspection of many of these has indicated we are adding additional sequence to many existing contigs (data not shown). More importantly, 1,123 sequence elements had no significant match in BtGI, but aligned with analogous, confirmed transcripts or genes in other animal species. A breakdown of databases to which these 1,123 elements matched is presented in Table 1. These sequences are being further analyzed to determine candidates to include on bovine oligonucleotide microarrays based on annotation. An additional 745 sequence elements were also identified as not matching any previously seen EST in current databases. To further characterize the quality





**Figure 1** Analysis of 8BOV sequencing efficiency: (A) sequencing redundancy analysis. The percentage of unique sequence (y-axis, left) being derived per plate (x-axis) was calculated throughout the sequencing. Cumulative unique and total sequence tallies are given on the right y-axis. (B) Post-clustering length distribution analysis. The total number of sequences (y-axis) comprising each sequence length range (x-axis) is shown.

of all sequence elements, we compared them to sequence information currently in the bovine trace archive at NCBI. Of 11,208 unique elements in this library, 11,034 of them (or 98.4%) match elements in the bovine DNA trace archive and therefore represent real, high-quality bovine sequences. Though it has not yet been determined what each of the unknown elements represent (indeed many may represent variants of known genes containing bovine introns, or untranslated regions of known genes), further analysis of these may also reveal novel bovine intestinal genes.

**Table 1** Summary of 8BOV library BLAST results

8BOV: 23,808 clones analyzed (62 384-well plates)

- 19,110 GenBank quality 5'-ESTs (80% success rate)  
Average sequence length: 600 bases
- 11,208 Unique elements: 7,789 singletons, 3,419 contigs  
[11,321 sequences]
- 11,034 (98.4%) match bovine DNA trace archive (NCBI)
- 9,340 Match *Bos taurus* Gene Index (BTGI) E  $10^{-10}$  (83%)
- 1,123 DID NOT match BtGI, but match other databases  
1,078 NR, non-redundant nucleotide (NCBI)  
1,073 Human genome database  
653 NR, non-redundant protein (NCBI)  
492 SwissProt protein database
- 745 Match no previously seen EST in any database

The normalized library (BARC-8BOV) contained  $1.8 \times 10^7$  cloned transformants with an average insert length of 1.86 kb as estimated by sizing on agarose gels PCR-generated products, or DNA fragments after restriction enzyme digestion, generated from over 200 clones (data not shown). Invitrogen reported a 20-fold normalization of the library by performing colony lifts to membranes and screening with a label for *elongation factor 1 alpha* (*EF1 $\alpha$* , a high copy number gene seen in other bovine libraries). Subsequently, EST from the non-normalized library (NN-8BOV) were surveyed. BLAST analysis was performed on 334 NN-8BOV clones, and the results from this analysis can be seen in Table 2. Interestingly, all except one of the most prevalent sequences in our 8BOV library (represented by 25 or more sequences each) were represented in the NN-8BOV non-normalized analysis (and estimated to be at a higher frequency) in this relatively small sampling. As expected, analysis of the composition of NN-8BOV showed a marked reduction in frequency of all NN-8BOV

**Table 2** Estimation of normalization procedure efficiency and reduction of most abundant sequences in BARC-8BOV

BARC-8BOV Seq.	BARC-8BOV Rank and identity	% in BARC-8BOV	% in NN-8BOV	EST fold reduction
134	1. IgJ Chain	0.70	1.8	2.6
106	2. Cytochrome B	0.55	1.5	2.7
65	3. Lysozyme	0.34	0.3	0
54	4. MHC ClassI HC	0.28	2.1	7.5
53	5. NADHHDH, Subunit 4	0.28	1.2	4.3
42	6. EF-1alpha	0.20	1.5	7.5
37	7. IgA Heavy Chain	0.19	1.2	6.0
30	8. IgM Heavy Chain	0.16	4.2	26.3
26	9. Beta-actin	0.14	0.9	6.4
146	Other 2-seq. contigs in NN-8BOV (14)	0.054*	0.6*	11.1

\*An average percentage was calculated.



**Table 3** Summary of functional annotation of all available 8BOV (intestinal) and 5BOV (mammary) normalized library sequence element GO terms into reduced level 3 GO annotation categories

GO term level 1	GO term level 2	GO term level 3	8BOV count	5BOV count
Biological process	biological process	biological process unknown	180	156
		cell communication	1611	1511
	cellular process	cellular physiological process	3122	3001
		cell differentiation	80	90
	development	development and growth	324	324
		morphogenesis	447	499
	physiological process	regulation of development	94	97
		death and aging	340	304
		metabolism	4327	3966
		organismal physiological process	861	843
		response to stimulus	990	938
	reg. of biological process	regulation of cellular process	548	557
		regulation of physiological process	829	731
	biological process	other processes	72	74
Cellular component	cellular component	cellular component unknown	222	222
		extracellular space/matrix	372	519
	cell	cell fraction/cell surface	377	331
		intracellular	3764	3673
		membrane	2036	1846

Molecular function	unlocalized	specific enzyme complexes	37	33
	molecular function	molecular function unknown	190	175
	binding	carbohydrate binding	36	31
		glycosaminoglycan binding	24	43
		protein and peptide binding	1223	1225
		nucleic acid and nucleotide binding	2169	1965
		lipid and steroid binding	112	93
		metal ion binding	701	604
		other binding activities	113	107
	catalytic activity	hydrolase and lyase activities	1744	1540
		kinase activity	520	466
		oxidoreductase activity	500	528
		transferase activity	1390	1225
		other catalytic and regulation activities	1073	1069
	transporter activity	protein/peptide transport	167	177
		lipid, carbohydrate, organic acid transport	123	91
		channel/pore class transport	97	68
		electron, ion transport	296	273
		carrier and other transport	486	504
	signal transducer activity	signal transduction processes	984	974
	transc./transl. reg. activity	transcription/translation regulation	658	566
	structural molecule activity	structural activities	300	365
		(cytoskeleton, ECM)		

contigs of three sequences or more. All other remaining two-sequence contigs in NN-8BOV also coded for well-represented genes/contigs in 8BOV. The average frequency in 8BOV of these two-sequence contigs was calculated, and an estimated 11-fold reduction was seen for these genes in 8BOV due to the normalization procedure. Of note, there were also two contigs present in NN-8BOV that have not yet been sequenced in 8BOV and may represent EST information that was lost because of the normalization procedure. In summary, the normalization procedure was effective and the degree to which it successfully removed high copy number sequences from the library was sequence-dependent and varied from 0 to over 20-fold. From our small sampling, we estimate an average 10-fold reduction for high copy number sequences in the 8BOV library.

To further characterize 8BOV, we summarized gene ontology terms associated with the 8BOV library as well as analogous terms from a previously described, normalized bovine mammary library, BARC-5BOV (6). Biological process, cellular distribution, and molecular function terms represented in 8BOV and 5BOV were scored and a summarization of the terms is given in Table 3. Each library was summarized in an identical manner, and about half of each libraries' unique sequence elements had available GO terms listed previously. The total GO terminology covered by both libraries is quite similar, demonstrating the general limitation of GO terminology in accurately defining overall library sequences. However, inspection shows 8BOV has a slightly larger representation of total genes involved in the GO level 2 term "cellular processes" (4,733 vs. 4,512) and in the level 3 categories of "metabolism" (4,327 vs. 3,966) and "regulation of physiological processes" (829 vs. 731). For 8BOV, there are a total of 1,611 in "cell communication," 3,122 sequences "in cellular and physiological processes," 861 in "organismal physiological processes," and 1,377 implicated in the regulation of cellular and physiological processes, all categories that were unchanged in the summarization and reduction of biological process terms from 48 to 14 categories. The mammary library 5BOV has similar representation in these categories and also a slightly higher representation of genes involved in the processes listed under the GO level 2 term of "development."

Likewise, an inspection of molecular functions (Table 3, bottom) demonstrates that a wide range of biochemical activities, such as binding, catalytic, or transporter activities, are represented in both the 8BOV and 5BOV libraries. For 8BOV, over 3,300 sequences code for proteins that have nucleic acid or protein binding capacities, suggesting involvement in regulatory processes at either the protein or DNA level, consistent with the need of these tissues to react quickly to different substrates provided (in blood or diet). While both libraries have similar representation in the molecular functions summarized, 8BOV also has a slightly higher representation of genes involved in binding activities, catalytic activities, and transporter activities related to metabolism. For 8BOV, catalytic activities (hydrolase/lyase—1,744 sequences; transferase—1,390 sequences; kinase/oxidoreductase—over 1,000 sequences; or other catalytic activities—over 1,000 sequences) are well represented, signifying we have successfully captured EST for a number of genes involved in nutrient uptake, metabolism, and assimilation. Of particular importance to both digestion and mammary gland function, both libraries contain a large number of genes with nutrient and ion transport activities, signal transduction activities, or that are involved in transcription or translation regulatory processes of the cell. Indeed, the similarities

between these two tissues are likely the result of their common physiological functions. Nutrient and ion transport are both functions of great importance to the lactating mammary gland and the intestine, one being a secretory organ absorbing nutrients from blood and exporting to milk, and the other an absorptive tissue capturing nutrients and transporting them to blood. In summary, we conclude that analogous to the mammary library, we have developed an intestinal library with a diverse representation of genes, involved in numerous biological processes of interest.

A summarization of all available GO annotation terms under the “biological process” main principle into major categories of functions that relate to metabolism was completed for 398 (35%) of the 1,123 putative physiologically relevant sequences in 8BOV not previously present in BtGI. The results of this summarization are shown in Table 4. The list of GO terminology for each sequence was assigned individually to group genes into major respective biological processes. By doing this, many sequences fit only under a single category, unlike the annotation from RefSeq where a single sequence may be represented in numerous similar or related GO categories. While a number of sequences were still assigned to more than one category because they had multiple functions and activities, absolutely no gene is represented in any one category more than once, and it was rare that any sequence had GO terms that caused it to be assigned under the same major heading (metabolism, transport, or major biological process) more than once. By this analysis, we demonstrate a number of these new bovine sequences encode for proteins involved in various aspects of ruminant metabolism and nutrient and ion transport processes. Sequences for over 150 genes involved in some form of metabolism, over 60 genes involved in transport, and hundreds of genes involved in major biological processes (cell cycle, signal transduction, etc.) have now been identified in cattle and have been added

**Table 4** Categorization of the novel, physiologically relevant 8BOV sequences having GO annotation into biological functional categories by scoring GO biological process terms

Biological process	Count
Biological process unknown	14
Metabolism	
carbohydrate metabolism	13
Protein metabolism	90
nucleic acid metabolism	35
lipid metabolism	17
Transport	
Protein transport	22
ion/e <sup>-</sup> /general transport	37
lipid and carbohydrate transport	5
Major biological processes	
cell cycle/cell growth	58
transcription/translation/repair/replication	84
apoptosis	14
signal transduction/cell-cell signaling	79
immune response	18
structural processes	8
other processes (i.e., adhesion, endocytosis)	60

to GenBank and to the *Bos taurus* Gene Index. Because these sequences with identifiable GO terms represent only one-third of the total number of bovine sequences matching transcripts in other animal species, these numbers actually represent only a fraction of our overall contribution. The BARC-8BOV library represents not only a large number of new bovine sequences, but also gene information for proteins involved in many important cellular processes.

In conclusion, BARC-8BOV is a valuable resource for researchers performing functional genomic studies of the bovine gut. With the advent of the complete bovine genome being available in the near future, it will also enhance efforts in detecting SNP, or other transcriptional variants that are representative of these gut tissues. Scientists at BARC have also recently developed an abomasal (9BOV) and intestinal (10BOV) library from tissues sampled after infection with gut nematodes *O. ostertagi* (abomasal library), and *C. oncophora* and *C. parvum* (intestinal library) (7). The sequence information generated from the BARC-8BOV library also complements these efforts to sequence and characterize digestive tract tissues from animals infected with nematodes. Together, these resources will be used to develop tools for functional genomic studies of the digestive tract. Currently, specific oligonucleotide arrays are being developed that contain sequences of metabolically important genes to assess the metabolic and physiological control of nutrient use by ruminant visceral tissues. Experimental analysis of gene expression, coupled with analysis of the proteome, will be essential to developing strategies for enhancing animal nutrient use efficiency and animal health. Putative targets for molecular-level manipulation via pharmacological or nutritional intervention will be clarified by these approaches.

## ACKNOWLEDGMENTS

The authors express their sincere gratitude for the expert technical assistance of Mary Niland, Tina Sphon, Joanne Wilson, and Larry Shade. Mention of trade names or commercial products in this paper is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.

## REFERENCES

1. Baldwin RL, McLeod KR. Effects of diet forage:concentrate ratio and metabolizable energy intake on isolated rumen epithelial cell metabolism in vitro. *J Anim Sci* 2000; 78:771–783.
2. Moody DE. Genomics techniques: An overview of methods for the study of gene expression. *J Anim Sci* 2001; (E. Suppl.):E128–E135.
3. van Hemert S, Ebbelaar BH, Smits MA, Rebel JM. Generation of EST and microarray resources for functional genomic studies on chicken intestinal health. *Anim Biotechnol* 2003; 14:133–143.
4. Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT, Casas E, Wray JE, White J, Cho J, Fahrenkrug SC, Bennett GL, Heaton MP, Laegreid WW, Rohrer GA, Chitko-McKown CG, Perteau G, Holt I, Karamycheva S, Liang F, Quackenbush J, Keele JW.

- Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res* 2001; 11:626–630.
5. Yao J, Burton JL, Saama P, Sipkovsky S, Coussens PM. Generation of EST and cDNA microarray resources for the study of bovine immunobiology. *Acta Vet Scand* 2001; 42:391–405.
  6. Sonstegard TS, Capuco AV, White J, Van Tassell CP, Connor EE, Cho J, Sultana R, Shade L, Wray JE, Wells KD, Quackenbush J. Analysis of bovine mammary gland EST and functional annotation of the *Bos taurus* gene index. *Mamm Genome* 2002; 13:373–379.
  7. Sonstegard TS, Van Tassell CP, Matukumalli LK, Harhay GP, Bosak S, Rubenfield M, Gasbarre LC. (Personal communication) regarding EST from cDNA libraries derived from immunologically activated bovine gut (manuscript in preparation, 2004).
  8. Baldwin RL, VI, McLeod KR, Klotz JL, Heitmann RN. Rumen development, intestinal growth, and hepatic metabolism in the pre and postweaning ruminant. *Dairy Sci* 2004; 87 (E. Suppl.):E55–E65.
  9. Smith TP, Godtel RA, Lee RT. PCR-based setup for high-throughput cDNA library sequencing on the ABI 3700 automated DNA sequencer. *Biotechniques* 2000; 29:698–700.
  10. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; 8:175–185.
  11. Matukumalli LK, Grefenstette JJ, Sonstegard TS, Van Tassell CP. EST-PAGE—Managing and analyzing EST data. *Bioinformatics* 2004; 20:286–288.
  12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res* 1997; 25:3389–3402.